Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal. Vol. 3, Issue 2 (2010), 126 – 133
ISSN 2070-5948, DOI 10.1285/i20705948v3n2p126

DIRECT CARDINAL INTERPOLATION

Mark R. Falknor, Eric M. Guild, Adam C. Hillier, Eric C. Like, and Steven C. Gustafson*

Air Force Institute of Technology, AFIT/ENG, Wright-Patterson AFB, Ohio 45433, United States of America

Received 31 March 2010; Accepted 11 July 2010
Available online 14 October 2010

Abstract: *Direct cardinal interpolation constructs a mean function that intersects given (x, y) points and a variance function that is zero at the points. These functions realize desirable extrapolation and efficiency properties for predicting y given x. It is found that direct cardinal interpolation is be more efficient than a classic form of Gaussian process interpolation in that its variance is typically much less over the point domain. It is also found that direct cardinal interpolation is less efficient near the end points (points not surrounded by other points); this desirable property is not realized by Gaussian process interpolation. These findings are a consequence of the direct construction of the mean and variance functions so that they achieve desirable properties.*

Keywords*: Interpolation, statistical efficiency, Gaussian process.*

1. Introduction

Given a set of (x, y) points generated by some process, interpolation, as considered here, specifies a mean function that intersects the points and a variance function that is zero at and only at the points. These functions have the following interpretation: if the process were to input the same x value and output a y value many times, then the mean function would give the mean of the y values and the variance function would give their variance. Thus interpolation generates a mean and a variance for the prediction of y given x, where for a point x value the point y value is predicted with zero error, i.e., the points are "noise free". Without loss of generality, the points $D= (x_i, y_i)$, $i = 1, 2, \ldots, n$, may be normalized so that their least squares line is the x axis and so

*Corresponding Author. Email: gustafson.steven@gmail.com

that they have zero mean and unit variance in both x and y. Thus the normalized points are such that $\Sigma x_i = 0$, $\Sigma y_i = 0$, $\Sigma x_i y_i = 0$, $\Sigma x_i^2 = n$, and $\Sigma y_i^2 = n$.

2. The classic linear model probability density

The classic linear model probability density $p(y|x, S)$ of y given x and the set of independent points $S = (x_i, y_i)$, $i = 1, 2, \ldots, n$ is derived by the following Bayesian analysis [4]. By marginalization:

$$p(y \mid x,\ S) = \int_{-\infty}^{+\infty} p(y \mid x,\ w)\ p(w \mid S)\ dw\ , \tag{1}$$

where w is the vector $(a\ b)T$. Let:

$$p(y \mid x, w)\ =\ (2\pi\sigma^2)^{-1/2} \exp[-(y - a - bx)^2 /(2\sigma^2)], \tag{2}$$

By Baye's rule:

$$p(w \mid S) = p(S \mid w)\ p(w) / \ p(S), \tag{3}$$

where from Equation (2):

$$p(S \mid w) = \prod_i p(y_i \mid x_i, w) = (2\pi\sigma^2)^{-1/2} \exp\left[-\sum_i (y_i - a - bx_i)^2 /(2\sigma^2) \right]. \tag{4}$$

Let the prior density $p(w)$ be uninformative in that it is constant. Then, Equations (1), (2), (3), and (4) show that:

$$p(y \mid x, S) = K \int_{-\infty}^{+\infty} \exp[-(y - a - bx)^2 /(2\sigma^2)] \exp[-\sum_i (y_i - a - bx_i)^2 /(2\sigma^2)]\ da\ db\ , \tag{5}$$

where K is a normalization constant. Analytical evaluation of Equation (5) shows that $p(y|x, S)$ is Gaussian with a mean that is the least squares line of the points and a variance that is the quadratic function $\sigma^2 + \sigma^2/n + (\sigma^2/\sigma_x^2)\ x^2/n$, where σ_x^2 is the variance of the point x values. For points that are normalized as indicated in Section 1, the quadratic function is $1 + 1/n + x^2/n$.

A classic constraint on the interpolation procedure is that the mean function and the variance function extrapolate to the mean and variance of the classic linear model density. As is intuitively reasonable, this constraint ensures that the least complex model, i.e., the linear model, applies for large magnitude x.

3. Gaussian process interpolation

A classic method for obtaining a mean function and a variance function is through the use of a Gaussian process [4, 5] or other methods, e.g., a spline technique. A Gaussian process supposes that the points S plus an additional point (x, y) sample a Gaussian probability density in n+1dimensions. If the process is zero mean, then it is specified by a covariance matrix:

$$H= h(x_j, x_k), \quad j, k= 1, 2, \ldots, n+1, \tag{6}$$

where $x_{n+1}= x$ and h is a given covariance function. Here H, which must be positive semi-definite and thus must have real and non-negative eigenvalues, may be partitioned [2] so that:

(1) The first n rows and columns form the matrix $\mathbf{M}$, which must also be positive semi-definite.
(2) The first n elements of the last column form the vector $\mathbf{v}$.
(3) The first n elements of the last row form $\mathbf{v}^T$, and the element in the last row and column form the scalar s.

Standard linear algebra identities [4, 5] then show that $p(y|x, S)$ is Gaussian with mean:

$$\mu(\mathbf{x})= \mathbf{v}^T\mathbf{M}^{-1}\mathbf{y} \tag{7}$$

and variance:

$$\sigma^2(\mathbf{x})= \mathbf{s} - \mathbf{v}^T\mathbf{M}^{-1}\mathbf{v}, \tag{8}$$

where $\mathbf{y}$ is the vector with elements y_i.

A classic form for a covariance function which requires that the mean function intersect the points and that the variance function be zero at the points is $\exp[-(x_j- x_k)^2/(2r^2)]$, where r is an adjustable parameter. If this form is multiplied by $(1+ 1/n + x^2/n)$, then, since v is zero for large magnitude x, the Gaussian process mean extrapolates to zero and the Gaussian process variance extrapolates to $s= 1+ 1/n+ x^2/n$, which for the normalized points is the variance of the classic linear model density. Note that for this classic covariance function, the Gaussian process mean and variance functions depend only on the width parameter r.

4. Direct cardinal interpolation

Direct cardinal interpolation explicitly constructs a mean function and a variance function that realize desirable extrapolation, smoothness, and efficiency properties. In the following, for two different mean functions, a classic meaning is that the smoothest function has the least roughness defined as integrated squared second derivative. Also, for two different variance functions, a classic meaning is that the most efficient function at a given x has the least variance.

For direct cardinal interpolation both the mean and the deviation functions have a classic form: a function which extrapolates linearly plus a weighted sum of Gaussian basis functions, each with its mean at a point x value and each with the same basis width. Here the deviation function is such that its square is the variance function, and thus deviation can be negative, unlike standard

deviation defined as the positive square root of variance. Deviation rather than variance is used since it extrapolates linearly and is not restricted to non-negative values.
The parameters in the form for the mean function are determined such that the function intersects the points, extrapolates to their least squares line, and has maximum smoothness. In particular, for direct cardinal interpolation applied to the normalized points, the mean function has the form:

$$\mu(x) = \Sigma A_i \exp[-(x - x_i)^2/(2s^2)] \quad (9)$$

Here the A_i are such that the points are interpolated, i.e., $\mu(x_i) = y_i$, and they are determined by solving n linear equations in n unknowns. The basis width s is then determined such that the mean function is maximally smooth, i.e., its roughness $\int_{-\infty}^{\infty} \mu(x)''^2\, dx$ is minimized.
The parameters in the form for the deviation function are determined such that the function is zero at the points, extrapolates to the deviation of the classic linear model, and has the minimum basis variance such that there are zeros only at the points. This minimum basis variance yields the minimum efficiency over the point domain (i.e., the x values between x_1 and x_n). Thus the deviation function is maximally "conservative". In particular, the variance function is $\sigma^2(x)$, where $\sigma(x)$ has the form:

$$\sigma(x) = (1 + 1/n + x^2/n)^{1/2} + \Sigma B_i \exp[-(x - x_i)^2/(2t^2)]. \quad (10)$$

Here the B_i are such that the variance is zero at the points, i.e., $\sigma(x_i) = 0$, and they are determined by solving n linear equations in n unknowns. The basis extent t (which is not the same as the basis width s in Equation 9) is then determined as the smallest value such that the variance function is zero only at the points. This determination is equivalent to minimizing efficiency or to maximizing the variance function in the point domain, i.e., t is determined so as to maximize $\sigma^2(x)$ for any $x_1 \leq x \leq x_n$. Note that for t smaller than this value the deviation function (and thus the variance function) has zeros not only at the x_i but also at other x. This outcome is not permitted since interpolation, as considered here, has zero variance only at the x_i.
Direct cardinal interpolation differs from cardinal interpolation [2], which finds the mean and variance functions using an infinite ensemble of basis functions, from robust cardinal interpolation [3], which requires each basis function in the infinite ensemble to be maximally smooth, and from fast cardinal interpolation [1], which uses the minimum possible number (four) of maximally smooth basis functions. Direct cardinal interpolation optimizes only two functions and is thus less computationally intensive. Most significantly, as the findings below indicate, direct cardinal interpolation can realize significantly greater efficiency over the point domain compared to a classic form of Gaussian process interpolation.

5. Comparison of direct cardinal and Gaussian process interpolation

The captions for four figures describe comparisons. Figure 1 compares direct cardinal interpolation with a classic form of Gaussian process interpolation for three normalized points. Figure 2 is similar to Figure 1 but for eight normalized points. Figure 3 compares the direct cardinal deviation function with the Gaussian process deviation function for the three points in Figure 1. Figure 4 is similar to Figure 3 but for the eight points in Figure 2. Note that maximum

likelihood (i.e., maximizing the probability of obtaining the given points as samples from the Gaussian density) is a classic method for determining the hyperparameter r for the Gaussian process. However, for comparison with the direct cardinal mean function, for which the basis width s is chosen such that the mean function has minimal roughness, the value of r is chosen such that the Gaussian process mean function has minimal roughness.

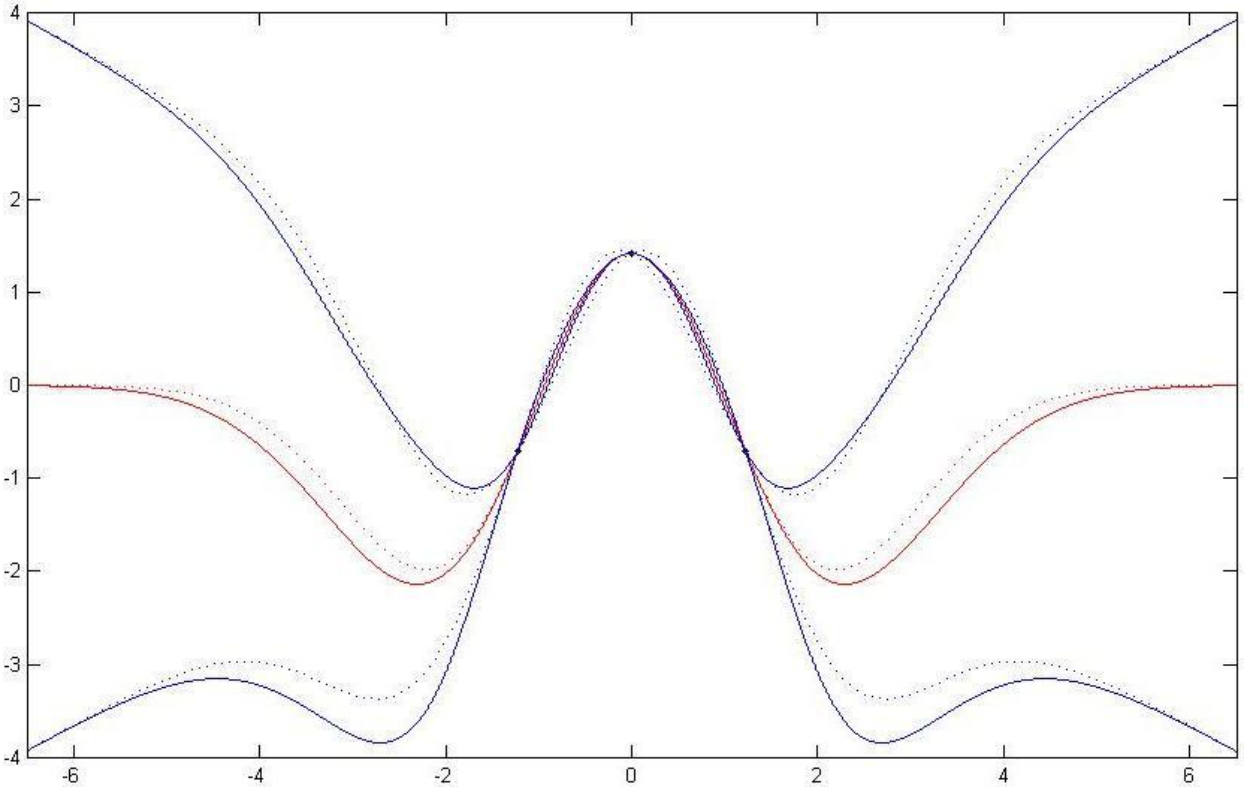

Figure 1. Shown are (1) three points normalized so that their least squares line is the x axis and so that they have zero mean and unit variance in both *x* and *y*, (2) the direct cardinal interpolation mean function (center solid curve), and this mean function plus and minus its deviation function (top and bottom solid curves), and (3) a classic Gaussian process mean function (center dotted curve) and this mean function plus and minus its deviation function (top and bottom dotted curves). As required for interpolation, both mean functions intersect the points and extrapolate to $y= 0$, and both deviation functions are zero at the points and extrapolate to $(4/3 + x^2/3)^{1/2}$. The Gaussian process mean function has a basis width of $r= 1.19$, which is the value that minimizes its roughness at 32.5. The direct cardinal mean function has a basis width of $s= 1.39$, which is the value that minimizes its roughness at 31.2, and the deviation function has a basis width of $t= 0.62$, which is the value that realizes the least efficiency for any *x* in the point domain (the range of x that encompasses the points). Note that both interpolation methods have minimally rough mean functions and are comparably smooth. However, the direct cardinal deviation function has significantly greater efficiency over the point domain than the Gaussian process deviation function. This greater efficiency is indicated by the significantly greater closeness of the direct cardinal deviation function to its mean function over the point domain compared to the Gaussian process deviation function.

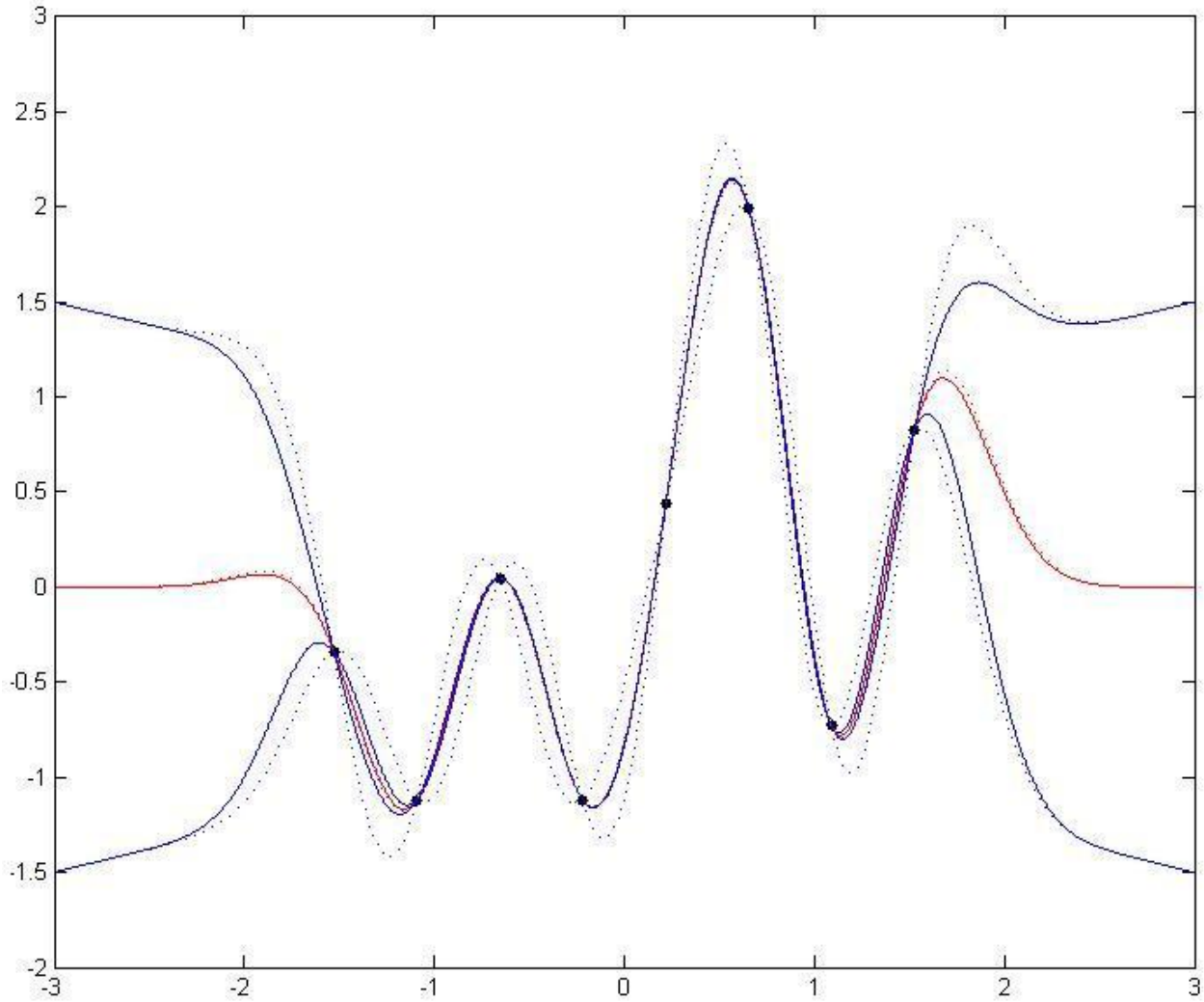

Figure 2. As in Figure 1, but for eight points with uniformly spaced x values and random y values normalized so that their least squares line is the x axis and so that they have zero mean and unit variance in both x and y. Here both deviation functions extrapolate to $(9/8 + x^2/8)^{1/2}$. The Gaussian process mean function has a basis width of r= 0.31, which is the value that minimizes its roughness at 1600. The direct cardinal mean function has a basis width of s= 0.31, which is the value that minimizes its roughness at 1610, and the deviation function has a basis width of t= 0.27, which is the value that realizes the least efficiency for any x in the point domain.

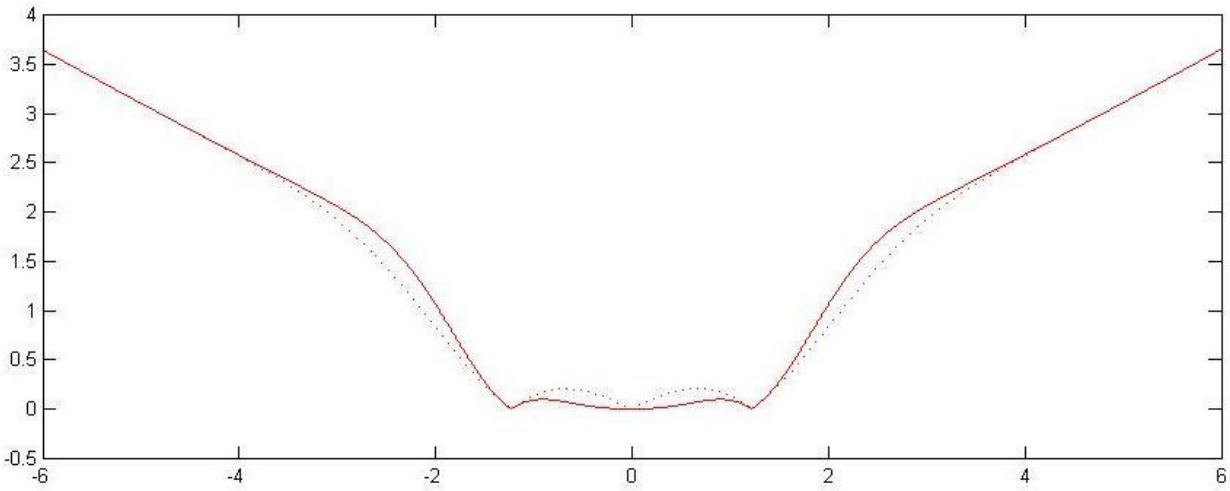

Figure 3. Shown for comparison are the direct cardinal deviation function (solid curve) and a classic Gaussian process deviation function (dotted curve) for the three points of Figure 1. Note that efficiency, as indicated by smaller variance and thus smaller deviation, is significantly greater over the point domain for direct cardinal interpolation. Note also that the direct cardinal deviation function, unlike the Gaussian process deviation function, indicates less efficiency near the end point *x* values than near other point *x* values---this result is intuitively reasonable.

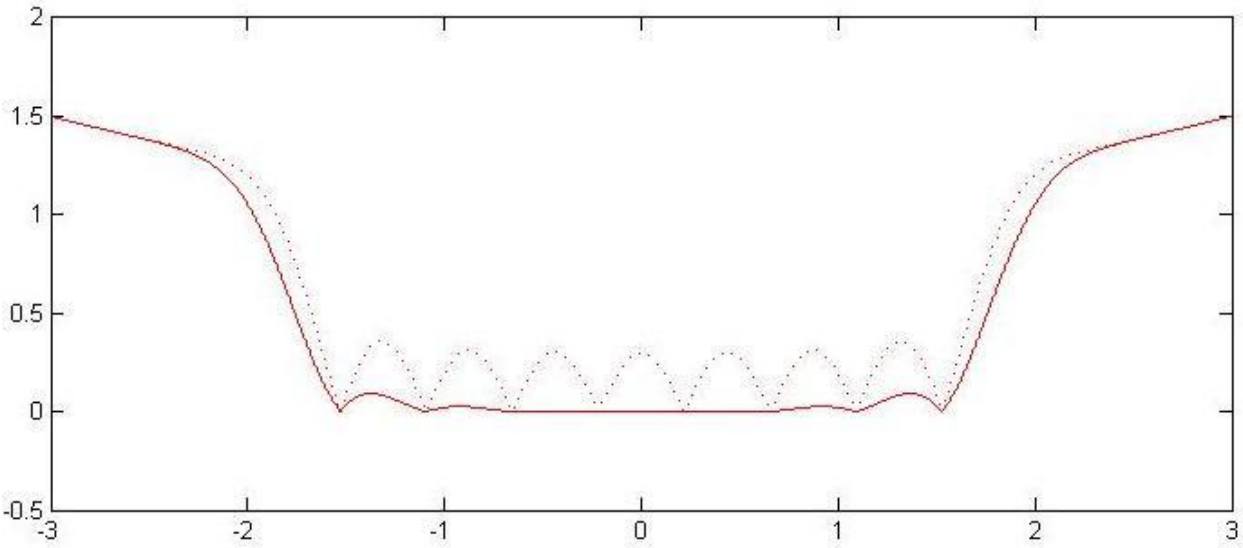

Figure 4. As in Figure 3, but for the eight points of Figure 2.

Acknowledgement

Helpful discussions with Richard K. Martin and Matthew C. Fickus are gratefully acknowledged. The following statement is required by the U.S. Government: The views expressed in this paper are those of the authors and do not reflect the official policies or positions of the United States Air Force, the US Department of Defense, or the US Government.

References

[1]. Guild, E.M., Like, E.C., Gustafson, S.C. (2009). Fast Cardinal Interpolation. *The Open Cybernetics and Systemics Journal*, 3, 85-89.
[2]. Gustafson, S.C., Parker, D.R., Martin, R.K. (2007). Cardinal Interpolation. *IEEE Transactions Pattern Analysis Machine Intelligence*, 29, 1538-1545.
[3]. Like, E.C., Gustafson, S.C. (2008). Robust Cardinal Interpolation. *The Open Cybernetics and Systemics Journal*, 2, 230 -233.
[4]. MacKay, D.J.C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge: Cambridge.
[5]. Rasmussen, C.,Williams, C.K.I. (2006). *Gaussian Processes for Machine Learning*. Boston: MIT Press.

Appendix

Theorem
There is a zero mean Gaussian process which generates any mean function $\mu(x)$ and any variance function $\sigma^2(x)$ chosen to model given points (x_i, y_i), where $i= 1,2,\ldots, n$ and $-\infty< x< \infty$.

Proof
As shown in Section 3 and in references [4, 5], a zero mean Gaussian process yields a mean function given by Equation (7) and a variance function give by Equation (8), where, as defined in Section 3, these equations are in terms of the matrix **M**, the vectors **v** and **y**, and the scalar *s* which partition the matrix **H**.
Without loss of generality, normalize the points so that $y_n= 1$. Let **M** be the *n* by *n* identity matrix, let **v** have zero elements except for $\mu(x)$ as the n^{th} element, and let *s* be $\mu^2(x)+ \sigma^2(x)$. Then Equations (7) and (8) are satisfied.
Furthermore, **M** has all unity eigenvalues and so is positive definite, as required. Finally, **H** has *n*-1 eigenvalues that are unity and two eigenvalues:

$$2^{-1}[1 + \mu^2(x) + \sigma^2(x)] \pm 2^{-1}\{[1 - \sigma^2(x) - \mu^2(x)]^2 + 4\mu^2(x)\}^{1/2},$$

which accordingly are real and non-negative, so **H** is positive semi-definite, as required. Thus **H** satisfies all requirements defined in Section 3, and accordingly there is a covariance matrix and thus a Gaussian process that generates any $\mu(x)$ and $\sigma^2(x)$. QED.

Remarks
Here **H** is obviously not unique. The form used above is "simple" in that only three of its elements are other than zero or one. However, it is "complicated" in that it is defined piecewise in terms of **M**, **v**, and *s*, and it does not correspond to any "classic" form. Note that normalization of the points so that $y_n= 1$ is not required if **H** is permitted to depend on y_n and y_n is not zero. Also note that additional normalization, such as $\Sigma x_i= 0$, $\Sigma y_i= 0$, $\Sigma x_i y_i= 0$, and $\Sigma x_i^2= n$, is permitted.